Science in China Series F: Information Sciences

www.scichina.com
info.scichina.com
www.springerlink.com

DRAGON-Lab —— Next generation internet technology experiment platform

WANG JiLong[1,2], LI ZhongHui[2], LÜ GuoHan[2], JIANG CaiPing[2], LI Xing[1,2] & ZHANG QianLi[2†]

[1] National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China;
[2] Network Research Center, Tsinghua University, Beijing 100084, China

Testbed technology is very important in the development of the Internet. Similar to the present internet, next generation internet also starts from testbed. There are two kinds of testbeds, testbed networks like CNGI-CERNET2, Internet2, Geant; testbed systems like PlanetLab, NS2. DRAGON-Lab can be viewed as both testbed network and testbed system. DRAGON-Lab is an independent AS (autonomous system) and connected to multiple real networks. On the other hand, DRAGON-Lab integrates many resources of its own, partners' and internet's, so as to provide open service. DRAGON-Lab has a large scale, provides open service, supports remote visualized experiments and programmable experiments. More details will be introduced in this paper.

Internet, testbed system, confederation

1 Introduction

Internet starts from testbed, which is always very important in internet development[1−3]. With the prevalence of the next generation internet research[4], the research on the next generation internet testbed platform is also becoming popular. In fact, the next generation internet research also started with testbed. NGI (next generation internet)[5] was first introduced in 1997 in USA. One of the first research activities was the testbed vBNS. Internet2[6], Geant[7], CNGI[8] were then developed by other nations in the following years.

Internet testbed platform can be classified into 2 categories: testbed network like CNGI-CERNET2[8], Internet2[6], Geant[7]; and testbed system like PlanetLab[9]. Testbed network is similar to real network; it is connected to the internet generally. Testbed network can be used to verify the current internet research and technology. The problem of testbed network is that normally testbed network is connected to the Internet and has a large population of users, thus mak-

Received June 3, 2008; accepted August 21, 2008
doi: 10.1007/s11432-008-0149-3
†Corresponding author (email: zhang@cernet.edu.cn)
Supported by the National Basic Research Development Program of China (Grant No. 2003CB314807) and the National High-Tech Research & Development Program of China (Grant No. 2005AA112130)

ing it difficult to test novel ideas. Often the testbed network is restricted to the member users. Also the testbed network has a high cost and strict rules for member users. On the other hand, it is comparatively easy to establish and make experiments on the testbed system. The testbed system provides an open simulation platform for small scale experiments. It is not a real network; thus the experiment is limited a little in scale.

Recently, USA and Europe have started the new internet testbed platforms GENI[10] and FIRE[11]. Both GENI and FIRE try to integrate the testbed network and testbed platform. For example, FIRE includes PanLab (Pan-European laboratory)[12] and OneLab (open networking laboratory)[13], which are aimed to set up a testbed confederation in Europe. The main purpose of PanLab is to set up a large testbed confederation so as to make the verification easy for new service, new concepts, and new architectures. Currently, PANLAB has 11 members. On the other hand, OneLab tries to focus on the development of PlanetLab in Europe. At present, PanLab and OneLab are still in the preliminary design status and there is still no clear idea on the resource integration and experiments design.

DRAGON-Lab is a confederation based large scale internet cooperative research platform. The confederation is a novel cooperating method. The two fundamental elements of confederation are autonomy and cooperation. Confederation member will contribute its resources to the confederation. Also the resource is maintained by its contributors. Confederation is a weak coupling cooperating method. The resources from confederation members may not provide service consistently. However, with a larger scale, the effects of random events may not affect the application considerably. The confederation also has a lower development cost.

The features of DRAGON-Lab can be summarized as follows:

● Confederation based model. DRAGON-Lab opens its network and hardware resources to the global researchers, aiming to cooperate with other testbed providers within the confederation model. It pays much attention to the internet free resources, and based on these resources DRAGON-Lab has built up the largest active measurement test system in the world.

● Specific purpose. DRAGON-Lab is not aimed to provide a silver bullet solution for all the internet experiments. It focuses on the support for traffic engineering research, measurement research and BGP related researches.

● Remotely visualized experiments configuration and environment build-up. DRAGON-Lab is a remote experiment environment; it uses a specific client to visualize the buildup of experiment environment.

● Programmable experiments. DRAGON-Lab proposes the programmable experiments idea. Based on this idea, the experiments environment can be realized by executable test scripts, thus making the test environment build up process automatic and repeatable.

More information about the DRAGON-Lab will be discussed in the subsequent chapters.

2 DRAGON-Lab architecture

2.1 DRAGON-Lab architecture

The architecture of DRAGON-Lab is shown in Figure 1. DRAGON-Lab has 4 layers: physical device layer, logical device layer, scheduling layer and service layer. The layer principle can be used to disguise the specifics of the lower layers, and provides a flexible and scalable interface. Physical device layer is constructed with all kinds of physical resources, like routers, switches,

servers and other measurement devices. Each physical resource has a corresponding DRAGON-Lab driver. Logical device layer includes all the DRAGON-Lab drivers and their DRAGON-Lab logical management interface. Users can access the resources via DRAGON-Lab drivers. Such architecture can simplify the integration of heterogeneous resources. To add a device into DRAGON-Lab, providing its corresponding DRAGON-Lab driver is enough. Scheduling layer is used to manage and schedule the experiments; it includes the validation of experiments, the generation of test scripts, the schedule of experiment time slot and the assignment of time slot and resources. In this process, all the operation is based on the logical device layer; researchers can build its test environment with a specific client. Test environment can be built with automatic DRAGON-Lab clients. For service layer, the experiment resources are encapsulated as a service. More service details will be given in section 3.

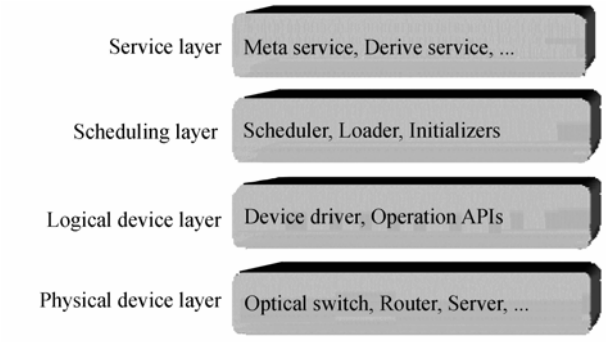

Figure 1 DRAGON-Lab architecture.

2.2 Experiment resources

DRAGON-Lab experiment resources include the following 3 categories:

1) Own resources.

The DRAGON-Lab's own resources are mainly from Tsinghua University and CERNET Center. These resources include:

● Network resources. DRAGON-Lab is a autonomous system of internet (AS24575), connected to CERNET, CERNET2, NSFCNET and TUNET with gigabits links.

● Network devices. DRAGON-Lab includes devices from Bitway, Huawei, ZTE, Cisco.

● Measurement devices. DRAGON-Lab has measurement device from IXIA, Agilent, Sperient and COMPASS.

● Servers. The servers in DRAGON-Lab are from IBM, LENOVO and DELL.

● Other devices. DRAGON-Lab also includes some devices designed by us. These devices support IPv4/IPv6, OSPF, BGP and multicast, and can be extended according to the specific requirements of the experiments. They can extend the services provided by the commercial products.

2) Partner resources.

● Measurement points. DRAGON-Lab has 61 measurement points across China. DRAGON-Lab also cooperates closely with other testbed organizations like FINET, NORDUNET, PlanetLab, PerfSonar. The cooperation is mutually beneficial.

● Real internet traffic and routing information. After a necessary de-privilege process, DRAGON-Lab provides traffic and routing information from TUNET and CERNET2. Both IPv4 and IPv6 resources are available.

● Network measurement and monitor tools. DRAGON-Lab is able to use the network measure and monitor tools developed by Tsinghua University. These tools can be used to monitor the experiment process.

3) Free service from the Internet.

DRAGON-Lab makes use of the free services from the Internet to provide more complicated services. For example, DRAGON-Lab uses the open service from the Internet to set up a distributed measurement platform, including about 1000 measurement points from more than 100 countries.

2.3 Switch engine

One of the DRAGON-Lab features is that the setup of test environment is automatically fulfilled by DRAGON-Lab. Switch engine is the component to implement this feature. It is located in the logical device layer. Switch engine will connect resources and automatically generate the environment. Switch engine is composed of an electronic switch engine and an optical switch engine. The architecture is shown in Figure 2.

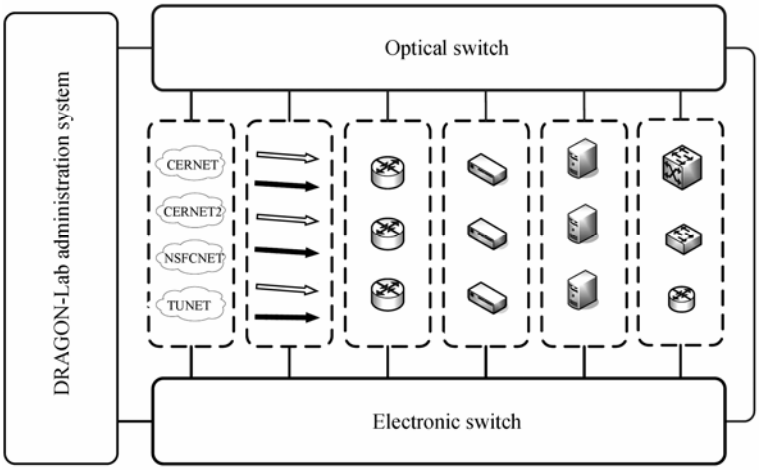

Figure 2 Switch engine.

Optical switch engine sets up the point-to-point link of devices. It provides the switch ability of layer 0—layer 1 and will not incur extra costs. It is used in traffic capture and device performance measurement experiments.

Electronic switch engine uses the Ethernet switch to provide layer 2 switch ability. It is used in scenarios where flexible topology or many interfaces are required. Since the switch engine itself may be the bottleneck of the throughput, it is not competent for device pressure test and throughput analysis experiments. Electronic switch engine can be the backup or extension of the optical switch engine.

2.4 Administration platform

DRAGON-Lab administration platform is made up of logical device layer identity and scheduling layer identity. The structure is shown in Figure 3.

The scheduler of DRAGON-Lab is based on the script language. The physical drivers are all described by the script language; even the experiment can also be defined as an executable script. The new device resource only requires a script driver to implement the predefined API.

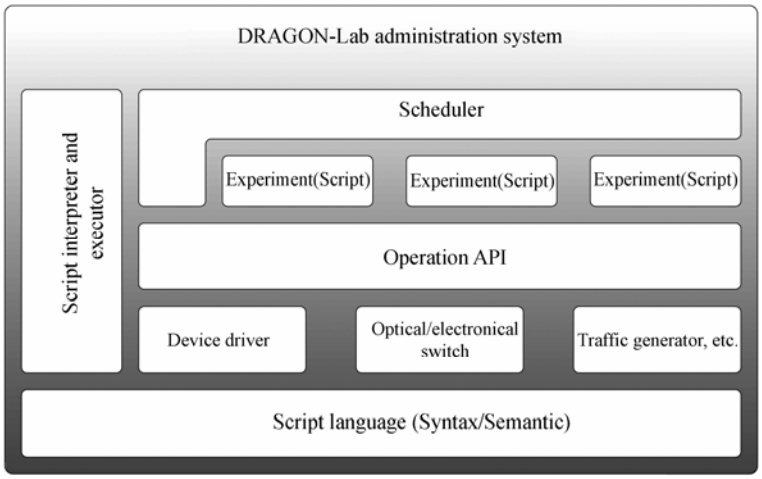

Figure 3 DRAGON-Lab administration platform.

The experiments are scheduled serially; one experiment can only be executed after the previous one. In the future we will consider the parallel execution of experiments.

2.5 DRAGON-Lab confederation

The main idea of DRAGON-Lab is to cooperate with other institutes and integrate resources from these institutes. This principle is also adopted by other platforms like PlanetLab.

The limitation of this cooperation is as follows:

● Lack of a cooperative method between platforms. For example, there are many well known distributed measurement platforms, but they are all operated independently and do not cooperate with each other.

● It is not easy to use the free service of the Internet. It is hard to grow up to a large scale only depending on the partners.

DRAGON-Lab invents the confederation architecture. DRAGON-Lab confederation is based on autonomy and cooperation. Members in the confederation will maintain the resources they contribute. There is no clear architecture of confederation; there are 3 forms of confederation: active confederation, passive confederation and mixed confederation, as shown in Figure 4.

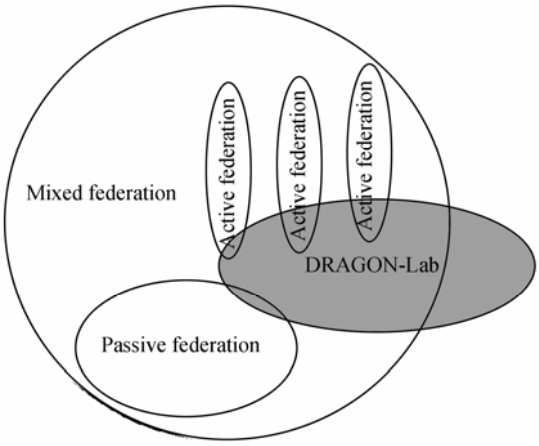

Figure 4 DRAGON-Lab confederation model.

● Active confederation is a confederation after negotiation. Now DRAGON-Lab has two methods to construct active confederation. One is to negotiate with every possible partner, and the other is to take part in an existing partner. A small resource may attain many partners.

● Passive confederation is to integrate free service from the Internet. There are many free services in the Internet. These free services can be freely used given no commercial commitment. Passive confederation is easy to develop and is very effective.

● Mixed confederation will integrate similar active confederation and passive confederation to

support specific experiment requirements. For example, DRAGON-Lab integrates many resources to build the largest globally distributed measurement system GPERF.

DRAGON-Lab uses the confederation model to extend the network resources, device resources and software resources.

3 Novel experiment technology

3.1 Remotely visualized experiment and programmable experiment

Internet experiments often encounter the following problems:

- Experiment geography problem. On-site experiments are often difficult to implement.
- Personal error. Internet experiment often requires the establishment of an experiment environment and device operation. These complex procedures may incur personal fault easily, especially for those operations with complicated operation and onerous device operation. The error is not easy to be fixed if it happens.
- Experiment repeating. It is often required to repeat an experiment, and thus the experiment environment should be reconstructed. The procedure may be very cumbersome if automatic environment setup is not possible.
- Efficiency problem. One experiment will last for a long period. For example, device testing may take more than 1 week even if there is previous experience.

DRAGON-Lab provides the remotely visualized experiment and programmable experiment.

1) Remote experiment environment visualization and automatic execution.

Researchers may use the DRAGON-Lab client to define the experiment. The environment can thus be generated automatically.

2) Programmable experiments.

The basic idea behind programmable experiments is to make the experiment procedure a script, thus avoiding unnecessary human intervention. For user requirements, DRAGON-Lab can interpret it as an executable script. This script can be imported, exported or executed. The execution will result in an experiment environment. The modification of script may result in a different experiment environment. To repeat the experiment, only the re-execution of the script is required. This feature makes it easy to repeat the re-implementation in the experiment.

3.2 Commodity hardware based on high speed traffic capture

High speed network traffic capture often requires dedicated-use devices. These devices are often very expensive and difficult to maintain. DRAGON-Lab integrates the technology of nCap[14] and PF_RING[15] to implement a zero-copy high speed commodity hardware based on traffic capture system in Linux system mpfring[16].

Mpfring provides two packet capture modes: one is single polling method and the other is dual polling method. Dual polling method implements the packets parallel processing. Although dual polling method is not efficient as single polling method, it provides more flexibility and can improve the performance in multi-CPU system. Thus the single polling method is used in scenarios where links number is larger than CPU number, while dual polling method is used when CPU number is much larger than links number.

3.3 High speed IP address anonymization scheme

When traffic data is shared by the partners, the IP address has to be anonymized. One of the most

important IP address anonymization schemes is prefix preserving IP address anonymization scheme. The prefix preserving IP address anonymization scheme requires that any two IP address with the same *k* bits prefix should be mapped to two IP addresses also with first *k* bits the same. To improve the prefix preserving IP address algorithm's performance, a series of fast prefix preserving IP address anonymization schemes are designed and implemented, including bit string based, block tree based, embedded bit string algorithm and the hybrid algorithm[17]. These algorithms are all 3—8 times faster than previous ones in making the real time anonymization of high speed network traffic possible.

4 Important services in DRAGON-Lab

DRAGON-Lab provides meta services and derived services. The meta services use the basic experiment resources in DRAGON-Lab, and the derived services use the meta services to provide more complicated ones. The derived services can be considered as special experiments in DRAGON-Lab, though the derived services can also provide some more advanced services for the users. Derived services can also be developed by the other partners.

4.1 Meta services

At present, DRAGON-Lab provides the following services:

1) Network experiment service. The network devices in DRAGON-Lab can be constructed into any testbed network. These networks can be used to implement routing, multicast, QoS, IPv6, wireless, and network management experiments. It is also possible to connect the testbed network to the Internet.

2) System experiment service. DRAGON-Lab enables remote login servers to implement network simulation experiments or application experiments.

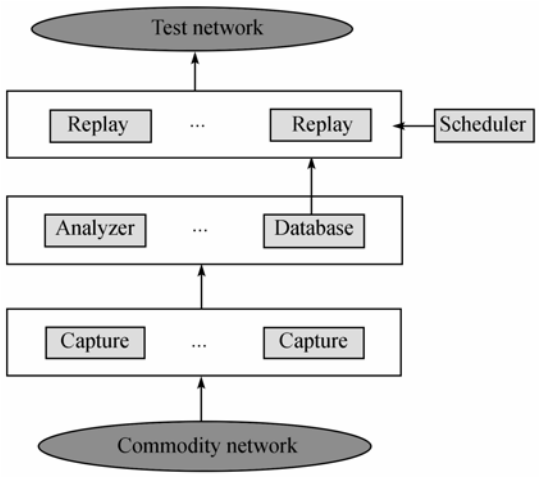

Figure 5 TRC architecture.

3) Real network traffic and routing information. The traffic data of TUNET and CERNET2 can be provided to the researchers after necessary de-privilege process. This traffic includes both IPv4 and IPv6 traffic data.

4.2 Traffic research center

Traffic research center (TRC) is a derived service. TRC can import traffic from real network and replay it if necessary. TRC can present the scenario in a real network. Its architecture is shown in Figure 5. The system includes 5 components: data capture system, data storage center, traffic replay system, control system and traffic analysis system. The basic idea of TRC is indicated in Figure 6.

4.3 Routing research confederation BGP- Grid

BGP-Grid is another important derived service in DRAGON-Lab, serving to support BGP routing research and experiment, as shown in Figures 7 and 8. The basic idea is:

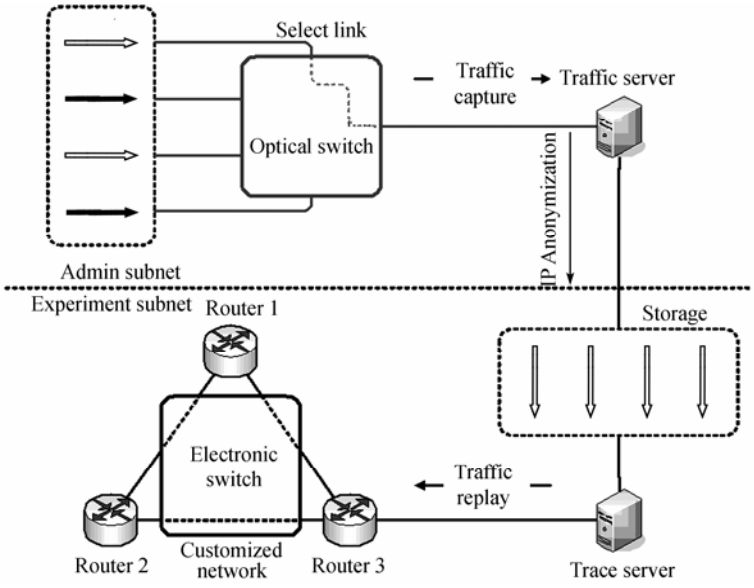

Figure 6 TRC implementation.

1) Each edge router will have a mirror device, which can learn the routing information and policy from the edge router.

2) According to the real topology among partner ASes, links between devices are set up using tunnel technology. Thus it is possible to simulate the routing process between ASes. Hence the policy optimization experiment can be implemented

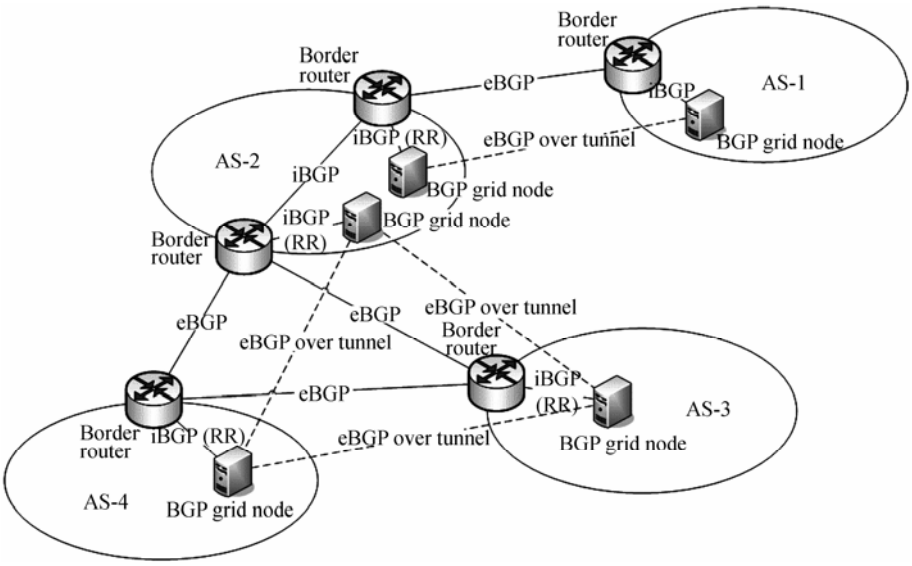

Figure 7 BGP-Grid.

The implementation of BGP-Grid requires the use of partner resources. At present, the partners include CERNET, CERNET2, TEIN2, NSFCNET, IPv6-CJ.

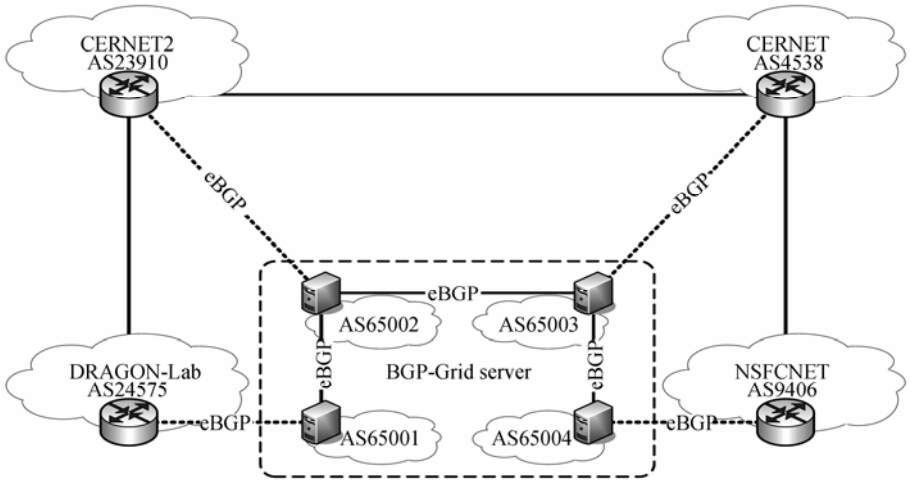

Figure 8. BGP-Grid implementation.

4.4 Global distributed performance measurement confederation GPERF

GPERF is a derived service for active measurement. The aim of GPERF is to support the world-wide active measurement. GPERF uses not only the services of DRAGON-Lab but also the partners' resources and free services from the Internet. At present, GPERF has about 1000 measurement points in the world; it is a mixed confederation system.

The design and implementation of GPERF are shown in Figures 9 and 10.

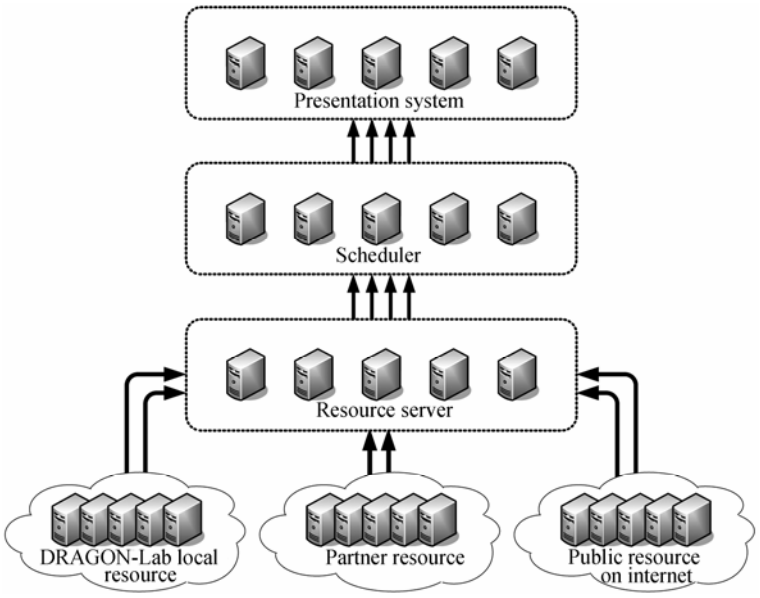

Figure 9 GPERF architecture.

5 DRAGON-Lab experiment process

To apply for resources and implement a service in DRGON-Lab testbed platform, five steps are required: experiment definition, experiment application, experiment schedule, experiment load, and experiment reset. The former two steps are finished by the applicant, while the latter three

steps are finished automatically by the testbed administration system.

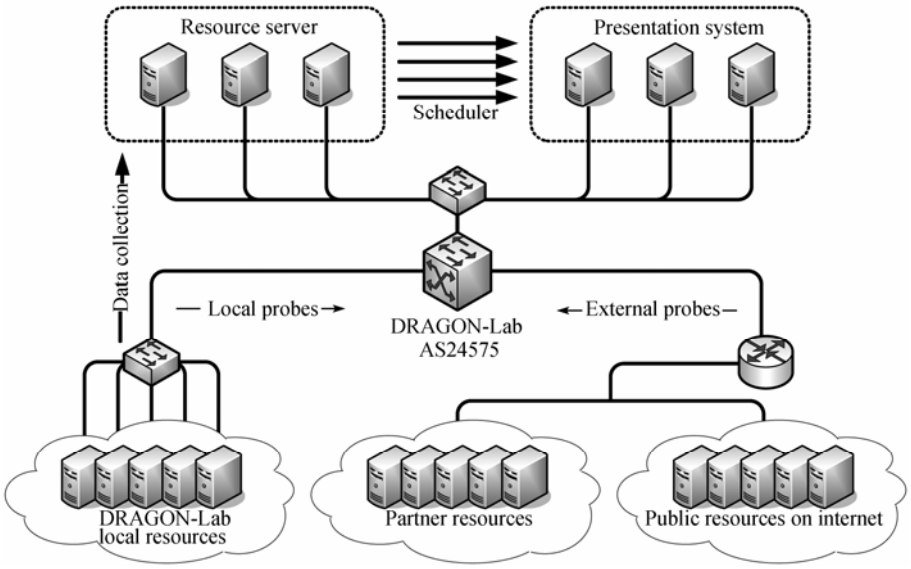

Figure 10 GPERF implementation.

Network devices, servers, special measurement device are often required in network related experiments. The experiment construction process can be viewed as an experiment topology design and then device port matching process. With methods similar to the traditional ones, researchers will finish the design in experiment definition step, which includes establishing the purpose, determining experiment steps, and deciding the topology. The difference between our method and the traditional methods is that the remote testbed platform based on experiment design does not require implementing the every step in experiment. With the DRAGON-Lab client software, it is only necessary to drag and drop the devices, design the topology visually, and then upload the design. The specific step from design to implementation is not necessarily implemented by the researcher; the backend software will set up the topology automatically. An experiment can be accomplished in six steps: new experiment definition, experiment device selection, ports selection, devices interconnection, saving the scheme, and uploading.

The predefined experiment is stored in testbed administration system as a script. Experiment will be assigned a time slot following his/her application. After the assignment of time slot, an email will be sent to the researchers titled "DRAGON-Lab time slot assignment" to notify the result. This process is the scheduler process.

The loading of experiment will be authorized by the platform administration system and then executed automatically and sequentially. The platform system will load the experiment automatically just before the authorized time slot. This process will schedule the experiment devices, configure the interconnecting ports, set up the environment and then send to applicant an email notification of the progress. The email will contain the information including the IP address of the environment, the login account and other necessary information. The platform system will send an email warning of the time slot end 24 hours before the time slot's end. DRAGON-Lab system will load the next experiment after the authorized time slot.

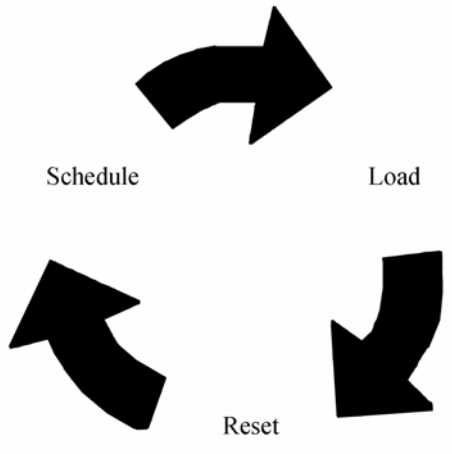

Figure 11 Experiment schedule, load and reset.

The experiment schedule, load and reset are indicated in Figure 11. Experiment is stored as a script. To reset the experiment, it is only necessary to reload the experiment.

6 Conclusions

Switch engines improve the automatic levels of DRAGON-Lab. Remotely visualized experiment technique improve the efficiency of DRAGON-Lab. Mpfring enables high speed network packet capture. High speed IP address anonymization scheme deprivilege the information to avoid the privacy concerns. The derived services like TRC, BGP-Grid, GPERF make DRAGON-Lab more practical. Layer architecture makes DRAGON-Lab open and scalable. Confederation based on cooperation model makes it possible for DRAGON-Lab to integrate many resources with low cost. DRAGON-Lab is still in an initial stage. Currently, it requires much effort to accomplish cooperating with more well-known international testbed platforms, integrating more free resources from internet, and improving in administration ability.

The authors thank Bin Tian, Jinpeng Jiang, Zhiyan Zheng and Linfeng Guan for their contribution to the DRAGON-Lab research.

1 Lu X C, Wang H M, Wang J. Internet-based virtual computing environment (iVCE): concepts and architecture. Sci China Ser F-Inf Sci, 2006, 49(6): 681—701
2 Mei H, Huang G, Zhao H Y, et al. A software architecture centric engineering approach for Internetware. Sci China Ser F-Inf Sci, 2006, 49(6): 702—730
3 Hou L S, Jin Z, Wu B D. Modeling and verifying Web services driven by requirements: an ontology-based approach. Sci China Ser F-Inf Sci, 2006, 49(6): 792—820
4 Wang J L, Ni C S, Wu J P, et al. Next generation internet. In: Proceedings of the International Symposium on Parallel Architectures, Algorithms and Networks. Washington: IEEE Computer Society, 2004. 351－356
5 Fowler D. The next internet. Networker, 1999, 3(3): 20－29
6 Internet2. http://www.internet2.edu/
7 Geant. http://www.geant.net/
8 CERNET2. http://www.cernet2.edu.cn/
9 Chun B, Culler D, Roscoe T, et al. PlanetLab: an overlay testbed for broad-coverage services. ACM SIGCOMM Comput Commun Rev, 2003, 33(3): 3－12
10 GENI Planning Group. GENI Design Principles. Computer, 2006, 39(9): 102－105
11 FIRE. http://www.future-internet.eu/
12 PanLab. http://www.panlab.net/
13 OneLab. http://www.fp6-ist-onelab.eu/wiki/view/OneLab
14 Deri L. nCap: wire-speed packet capture and transmission. In: Proceedings of the End-to-End Monitoring Techniques and Services. Washington: IEEE, 2005. 47－55
15 Deri L. Improving passive packet capture: beyond device polling. In: Proc SANE, 2004
16 Shen W C, Yang B, Lv G H, et al. Commodity PC based on flow aggregation system for gigabit networks. J Xiamen Univ, 2007, z2
17 Zhang Q L, Wang J L, Li X. On the design of fast prefix-preserving IP address anonymization scheme. LNCS 4861. In: Proc ICICS. Heidelberg: Springer, 2007